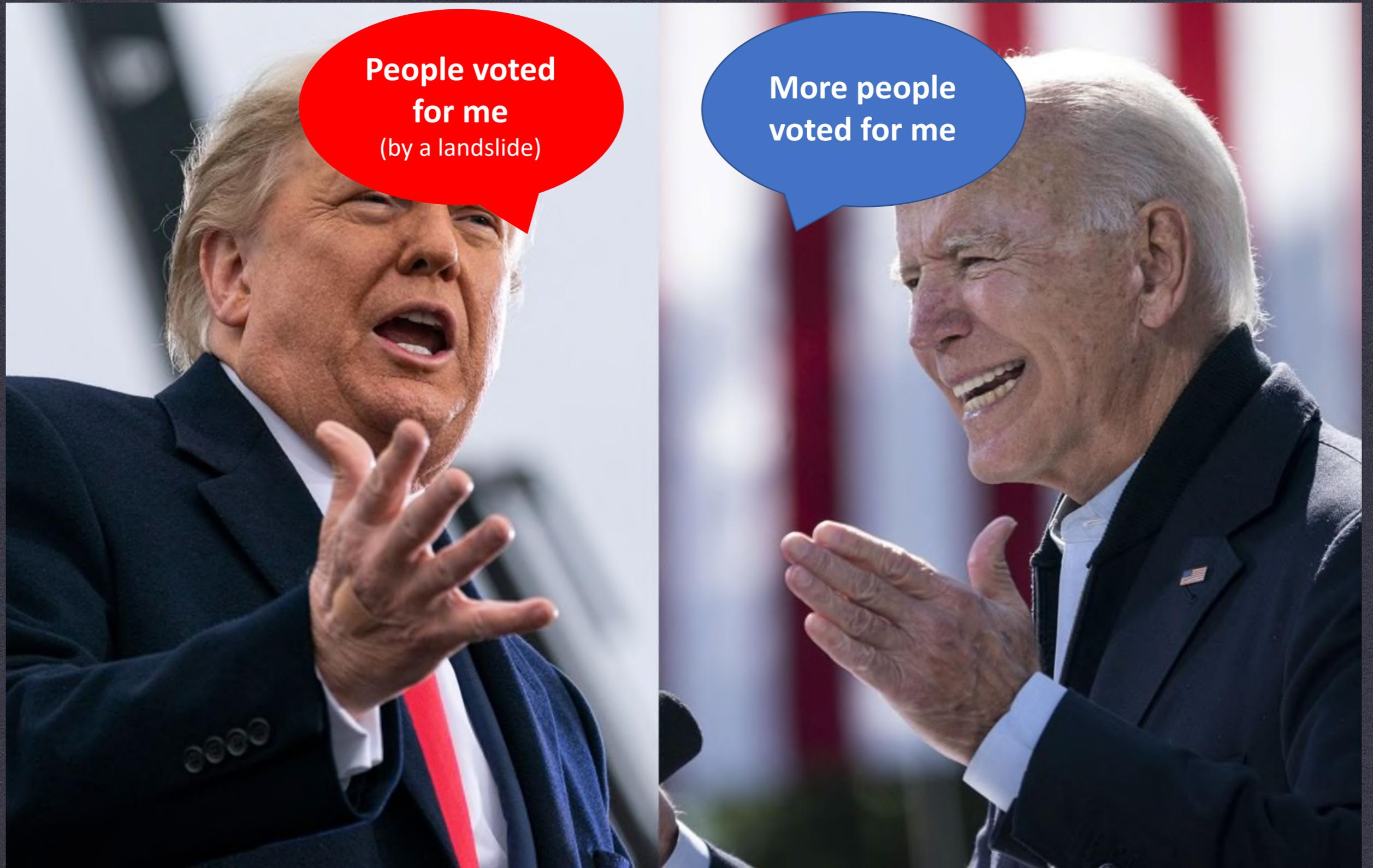




# **VERS LA NÉCESSITÉ D'ÉTUDIER LES LIENS ENTRE LES VARIABLES**

**QUELQUES EXEMPLES**



People voted  
for me  
(by a landslide)

More people  
voted for me

# LES ÉLECTIONS PRÉSIDENTIELLES AMÉRICAINES

# Les élections présidentielles américaines

- \* Qui a voté pour quel candidat ? Une questions très importante en sociologie
- \* On cherche à déterminer le profil des individus ayant voté pour *Donald Trump* de ceux qui votent pour *Joe Biden*
- \* Cela revient à déterminer les caractéristiques qui permettent de déterminer si une personne va voter pour le parti Républicain ou plutôt pour le parti Démocrate.

**On peut comparer cela à la consommation de thé glacé**

**Qui sont les consommateurs de la boisson May Tea ?**

**Qui sont les consommateurs de la boisson Lipton ?**

**On peut faire le même travail que pour les Élections Présidentielles Américaines**



**LES CONSOMMATEURS DE THÉ GLACÉ**

# Concernant les deux exemples

- \* On commence par s'intéresser à la nature des variables : choix du candidat au moment du vote ? Choix de la boisson que l'on va consommer ?

## **Une variable qualitative**

- \* Pour cette étude, on est amené à se baser sur des facteurs sociaux-démographiques : catégorie âge, statut social, statut professionnel, lieu de résidence, ethnicité, appréciation du prix du produit ...

## ***Une autre variable qualitative***

- \* Cela doit vous rappeler un test statistique abordé lors d'un précédent cours ... on en reparlera par la suite
- \* Un dernier exemple serait de savoir si une nouvelle procédure mise en place dans une entreprise permet d'augmenter la productivité/performance de ses salariés

## Comparer le prix d'une pizza

On cherche à savoir si le prix d'une pizza varie en fonction de la localité.

Les pizzas parisiennes sont-elles plus chères que les pizzas de Bar-le-Duc ?

On relève le prix des pizzas dans les régions pour effectuer notre analyse

Regardons encore un autre exemple



**PRIX D'UNE PIZZA V.S. LOCALITÉ**



## **LE CAS DU CORONAVIRUS - TAUX D'INCIDENCE**

ON PEUT L'ÉtudIER SUIVANT PLUSIEURS CRITÈRES SOCIAUX DÉMOGRAPHIQUES

# Coronavirus

- \* L'objet de l'étude est ici le taux d'incidence dans la population

## Une variable quantitative

- \* Une étude que l'on peut mener selon différents critères comme *l'âge de la population, lieu de résidence, régimes alimentaire, pratique ou non d'une activité sportive, antécédents médicaux, ...*

## Une variable qualitative

- \* Cela nous conduit à effectuer un test corrélation dont la nature sera différente du test précédent entre deux variables qualitatives



## **NOTE D'UN EXAMEN VS TEMPS PASSÉ À RÉVISER**

**EST-CE QUE LE TEMPS PASSER À ÉTUDIER A UNE INFLUENCE SUR VOTRE NOTE À UN EXAMEN ?**

# Un autre exemple

- \* Est-ce que la puissance d'une voiture est proportionnelle à sa cylindrée ?
- \* Est-ce que le temps de passage du 0 à 100 km/h est proportionnelle à la puissance du véhicule ?



# Notes et Puissance

- \* L'objet de l'étude la note obtenue à un examen ou encore la puissance d'une voiture

**Une variable quantitative**

- \* Cette étude peut-être menée selon plusieurs critères :

1. Le *QI de l'étudiant* ou encore le *temps passé à réviser*. Mais on peut encore s'intéresser à des critères plus fous comme le *temps que l'étudiant à passer à composer*
2. La *cylindrée* ou la *puissance du véhicule*. On pourrait également regarder s'il y a un lien avec le *prix du véhicule*.

**Une variable quantitative**

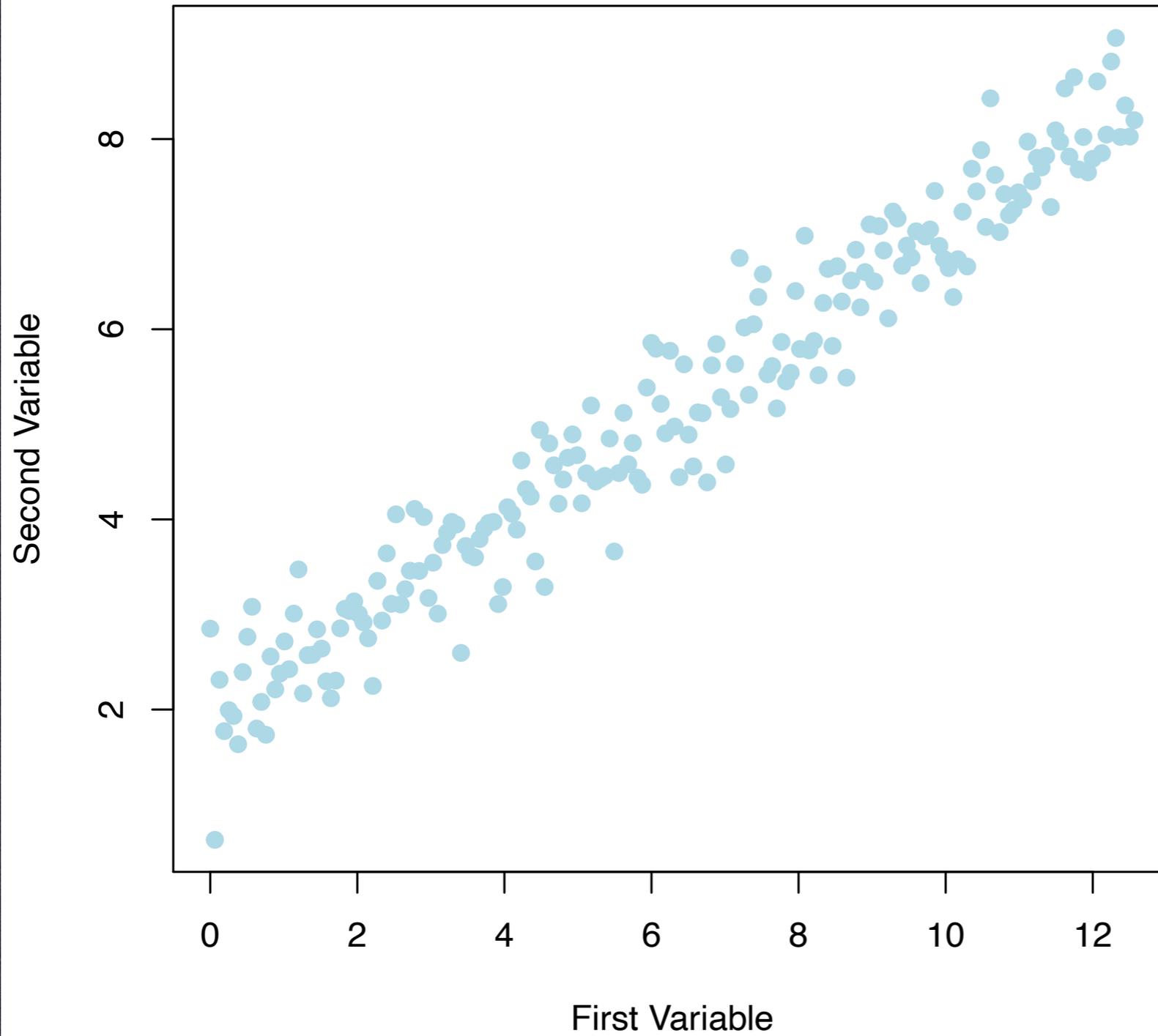
- \* Cela nous conduit à effectuer un test corrélation dont la nature sera différente du test précédent : corrélation entre deux variables quantitatives

**Regardons la notion de corrélation entre deux variables quantitatives d'un peu plus près**

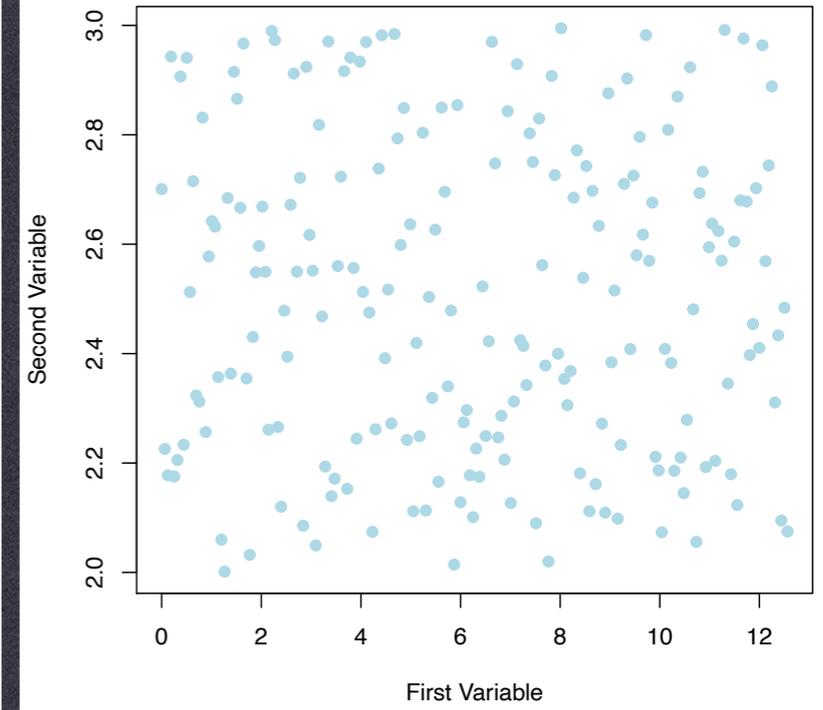
# **CORRÉLATIONS**

**QUELQUES DÉFINITIONS DANS LE CAS DE DEUX  
VARIABLES QUANTITATIVES**

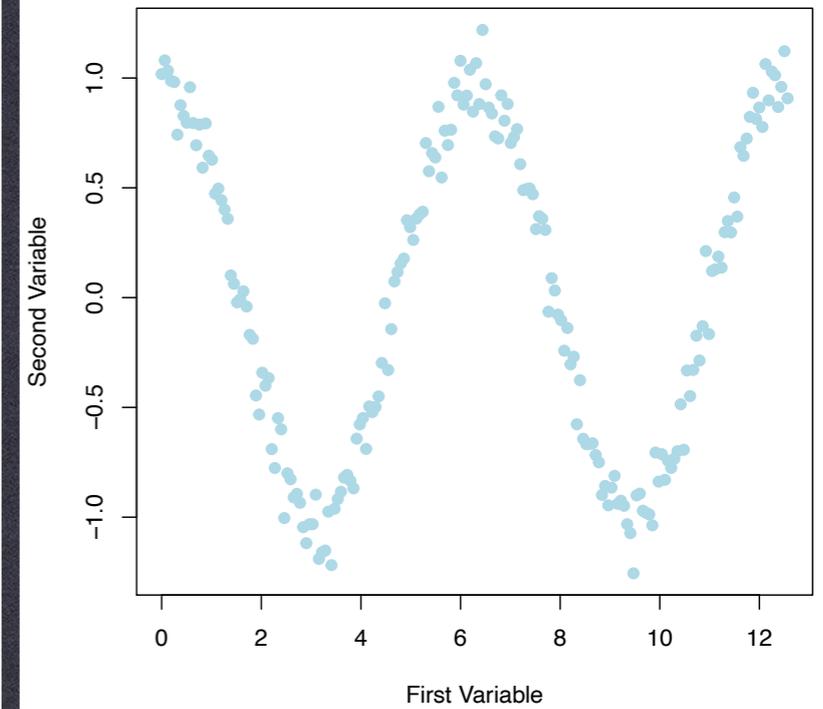
## Linear Correlation



## No Correlation



## Non Linear Correlation



# ETUDE DES CORRÉLATIONS

QUELQUES EXEMPLES DE CORRÉLATIONS OU D'ABSENCES DE CORRÉLATIONS

# Corrélations

- \* L'objectif est de savoir si deux variables quantitatives sont corrélées ou non mais également de savoir si la liaison est forte !  
Le cas des variables quantitatives car nous sommes capables de caractériser
  1. la **présence ou non** d'une corrélation entre des variables
  2. l'**intensité** de la liaison entre les deux variables (quand la liaison entre les deux variables est *linéaire uniquement*)
- \* Le premier point se fait à l'aide d'un test statistique reposant sur la *loi de Student*. Pour quantifier l'intensité entre deux variables, on calcule ce que l'on appelle **le coefficient de corrélation de Pearson (1857-1927)**

# Corrélations : définition

- \* Le coefficient de corrélation de Pearson permet de déterminer comment la variable 1 agit sur la variable 2
- \* Formellement, si on considère deux variables aléatoires  $X$  et  $Y$ , alors ce coefficient, noté  $\rho$  est défini par :

$$\rho = \frac{(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

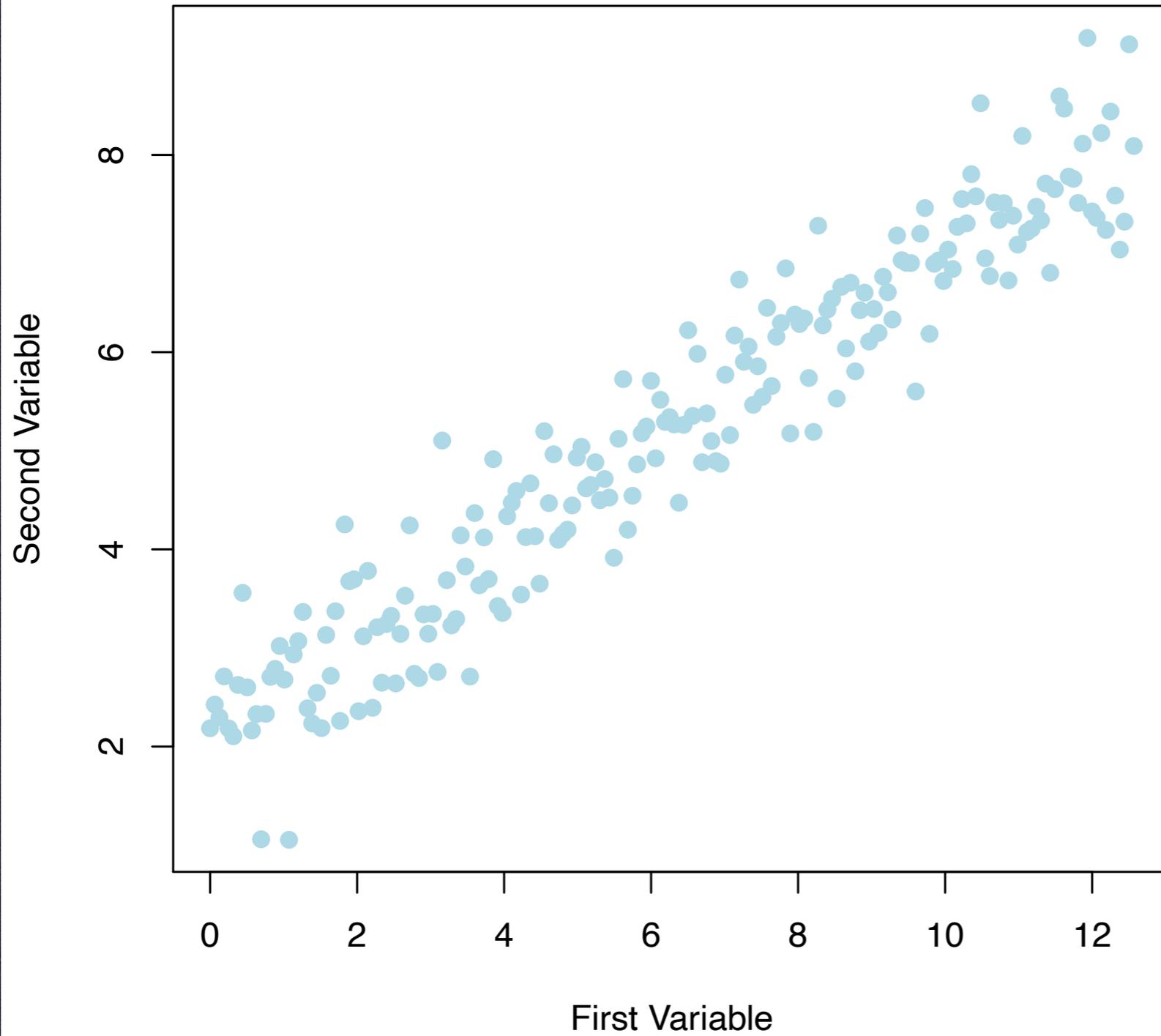
- \* Si l'on dispose de deux échantillons  $(y_i)_{i=1}^m$  et  $(x_i)_{i=1}^m$  on peut déterminer une valeur empirique de ce coefficient de corrélation  $\hat{\rho}$  définie par

$$\hat{\rho} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}}$$

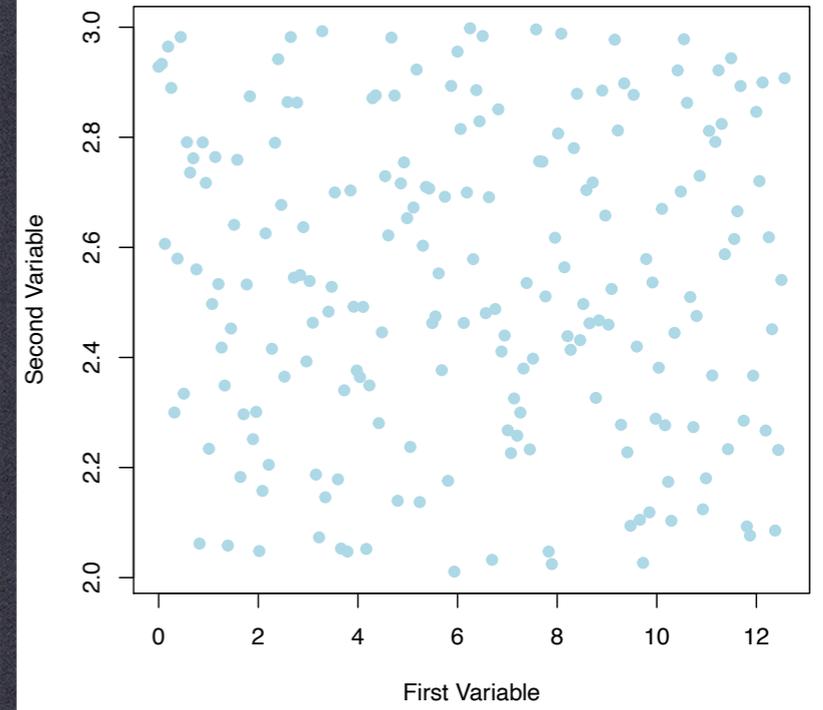
# Corrélations

- \* Le coefficient de corrélation prend ses valeurs dans  $[-1,1]$
- \* **Interprétation** : selon la valeur du coefficient, les conclusions varient selon sa valeur dans cet intervalle,
  - si sa valeur est **négative** (proche de -1) on dit que **la corrélation est négative**, *i.e.* des **valeurs croissantes** de la première variable impliquent des **valeurs décroissantes** de la variable 2
  - si la valeur est **positive** (proche de 1), **la corrélation est alors positive**, des **valeurs croissantes** de la variable 1 impliquent des **valeurs croissantes** de la variable 2
  - dans le cas où la valeur est **proche de 0**, on dit que les variables sont **faiblement voire non corrélées**.

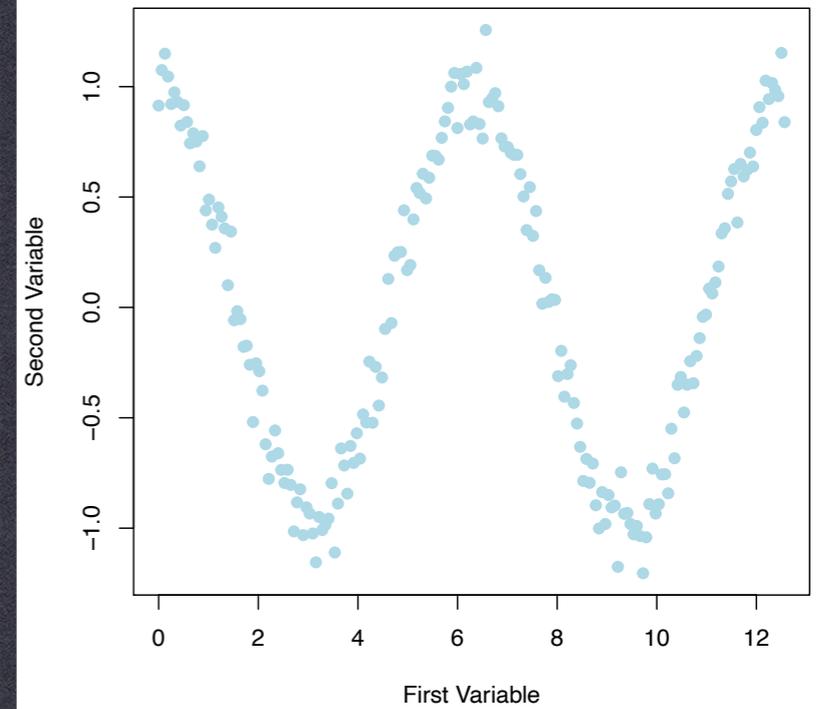
**Linear Correlation,  $r = 0.958$**



**No Correlation,  $r = -0.037$**

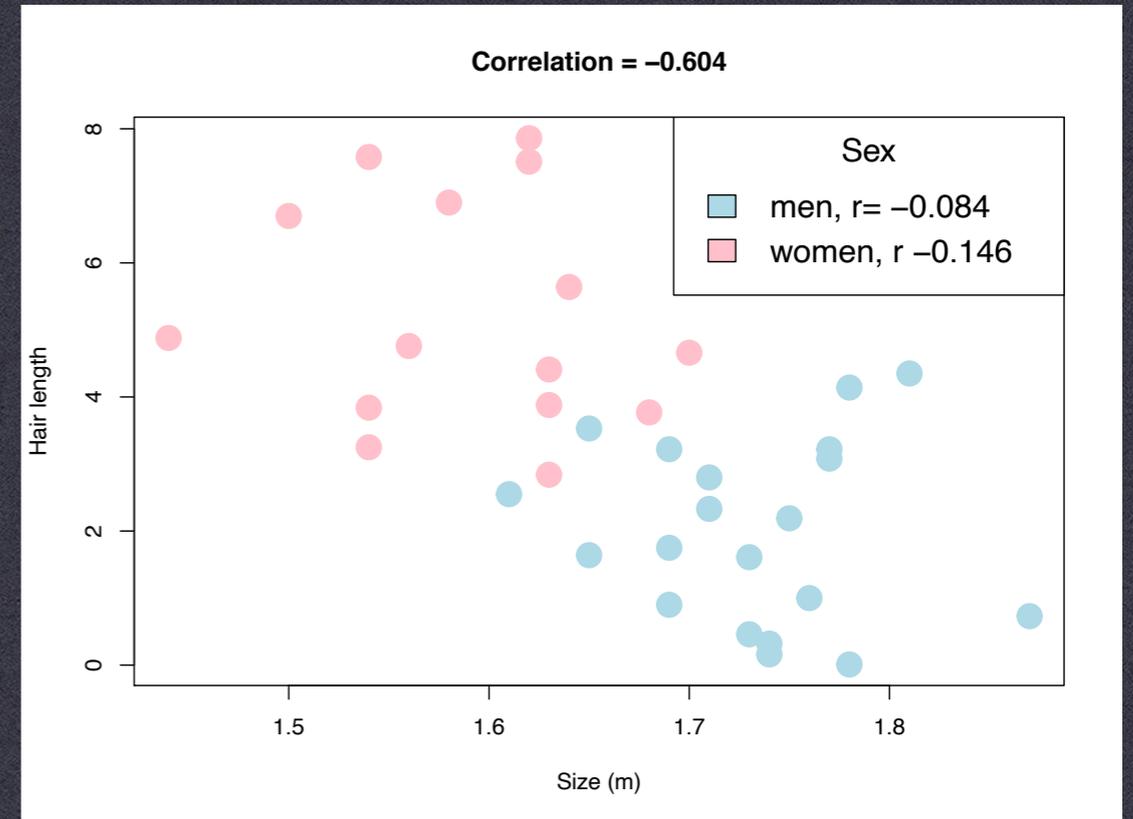
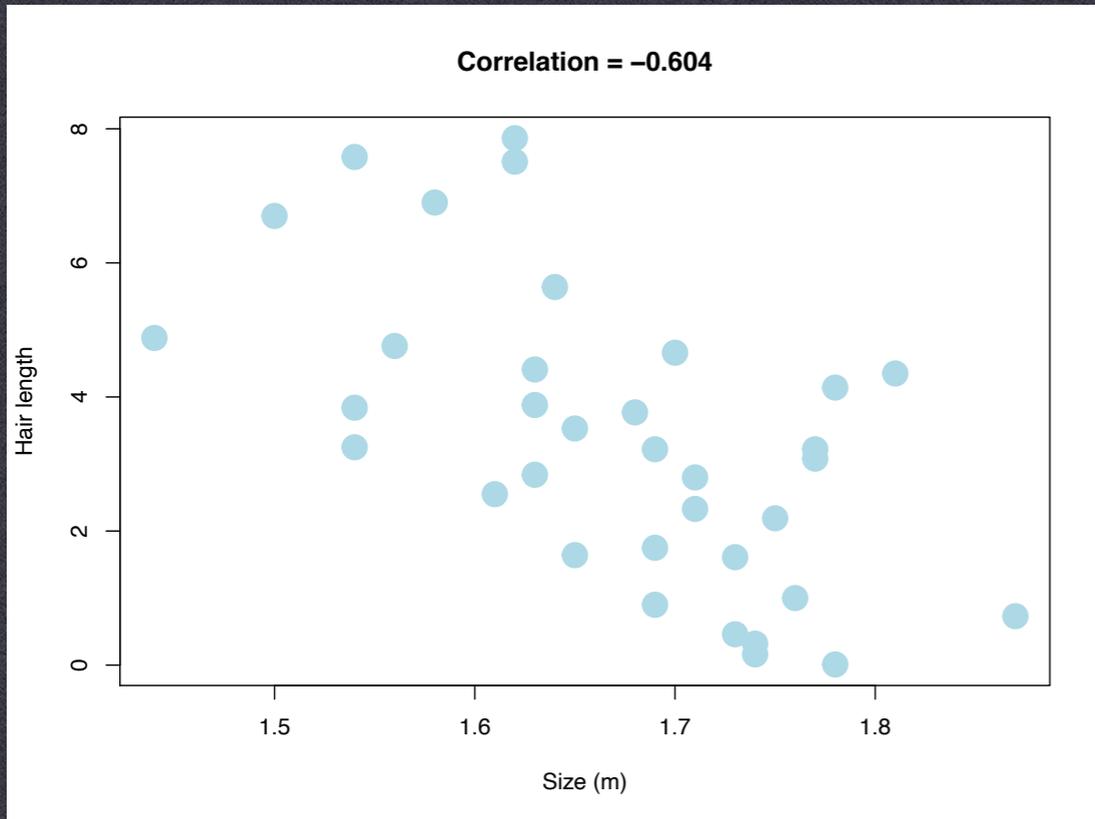


**Non Linear Correlation,  $r = -0.014$**



# ETUDE DES CORRÉLATIONS

CALCUL DU COEFFICIENT DE CORRÉLATION



**Méfiez-vous des corrélations trop « évidentes »**

**Certains facteurs peuvent expliquer une corrélation et la prise en compte de ce facteur permet d'effacer toute corrélation précédemment établie.**

**ETUDE DES CORRÉLATIONS : TAILLE INDIVIDUS VS LONGUEUR CHEVEUX**  
**CORRÉLATION ET FACTEURS SOUS-JACENTS**

# Corrélation : un exemple

- \* On considère le tableau de données suivantes et on vous demande de regarder si les deux variables sont corrélées ou non.

|   |    |    |      |      |     |      |      |      |      |      |
|---|----|----|------|------|-----|------|------|------|------|------|
| x | 9  | 10 | 16   | 12.5 | 19  | 22.5 | 16   | 12.5 | 17   | 28   |
| y | 14 | 12 | 11.1 | 14.2 | 9.7 | 7.5  | 11.1 | 11.4 | 11.8 | 10.1 |

1. Calculer le coefficient de corrélation de Pearson
2. En utilisant la statistique de test définie ci-dessus et le fait qu'elle suive une loi de Student à  $n-2$  degrés de liberté, peut-on dire que la corrélation est significative

$$T = \frac{\hat{\rho}}{\sqrt{\frac{1 - \hat{\rho}^2}{n - 2}}}$$

# THÉORIE DES TESTS

QUELQUES RAPPELS RAPIDES

# Théorie des tests

- \* On fixe une hypothèse nulle  $H_0$  que l'on considère vraie : **les variables étudiées sont indépendantes**
- \* On se fixe un seuil de risque de première espèce  $\alpha$  (probabilité de rejeter à tort)
- \* On calcule la valeur de la statistique de test
- \* On en déduit alors la p-value à partir de la loi de la loi de la Statistique de Test
- \* On compare la p-value au risque de première espèce  $\alpha$  . Si la **p-value est inférieure au risque de première espèce  $\alpha$** , on rejette alors l'hypothèse et on conclut à la **corrélacion entre les deux variables**. Dans le cas contraire, les données ne permettent pas de conclure au rejet de l'hypothèse nulle.

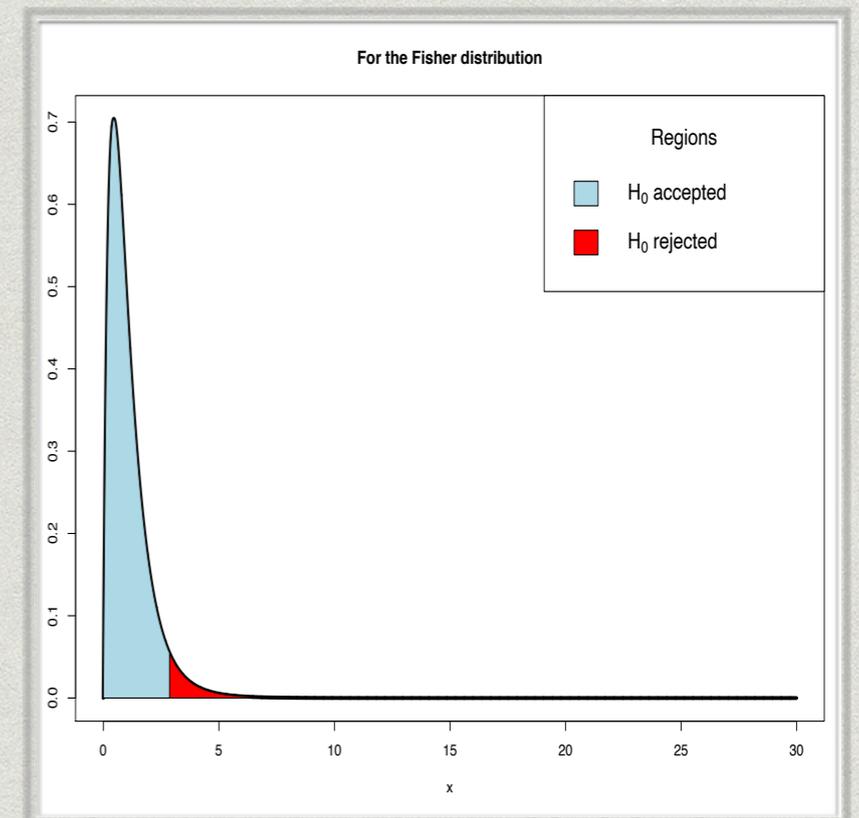
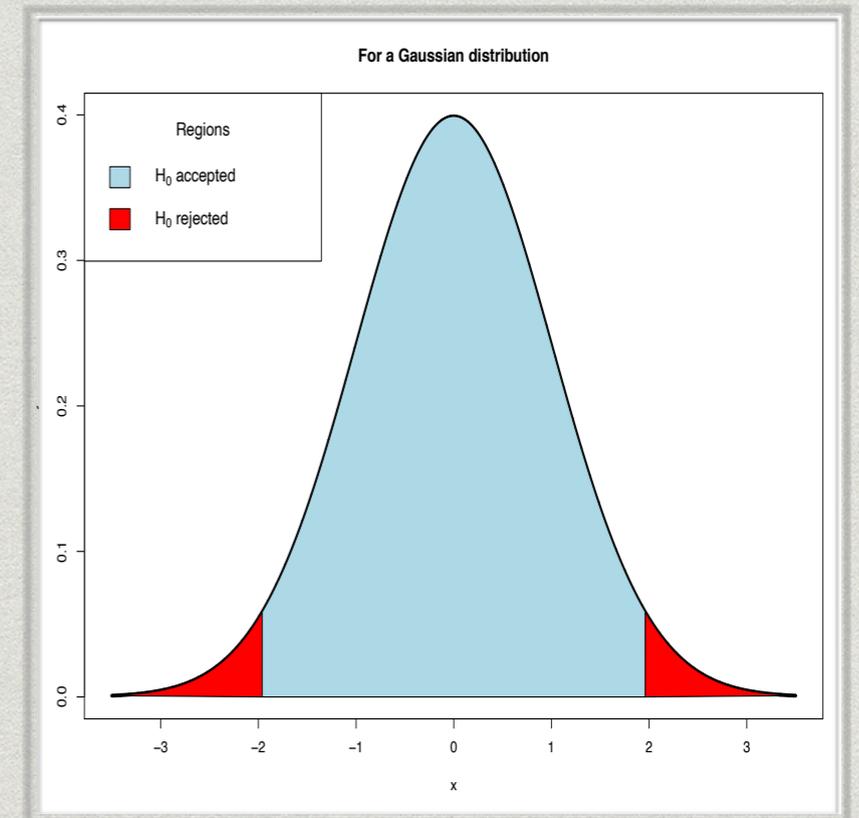
# Théorie des tests

\* Deux grandes catégories de tests :

1. Les tests bilatéraux

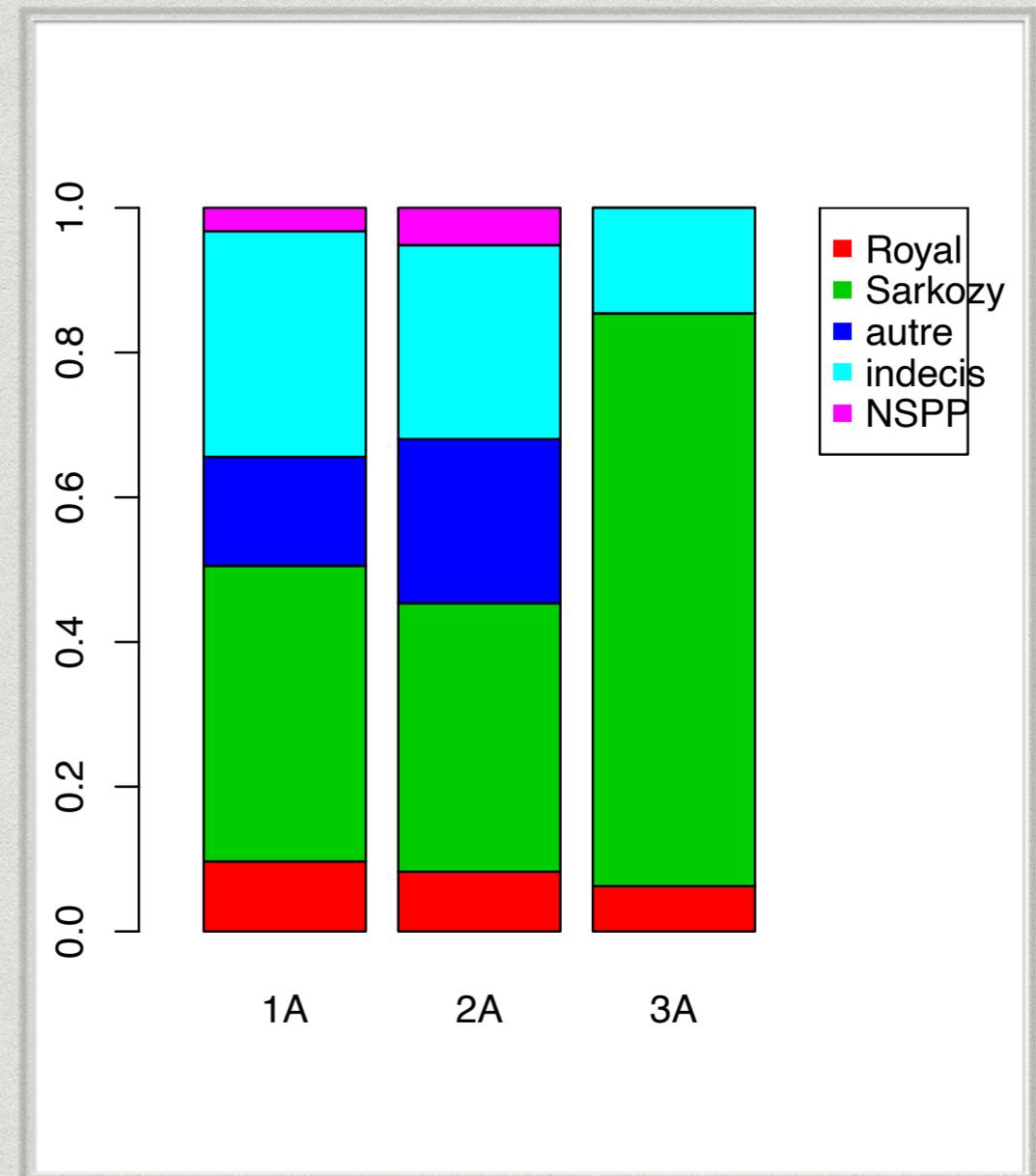
2. Les tests unilatéraux

\* A chaque valeur critique est associée une p-value (aire en rouge sous la densité). Définie les valeurs de la statistique de test pour lesquelles on va rejeter l'hypothèse nulle.



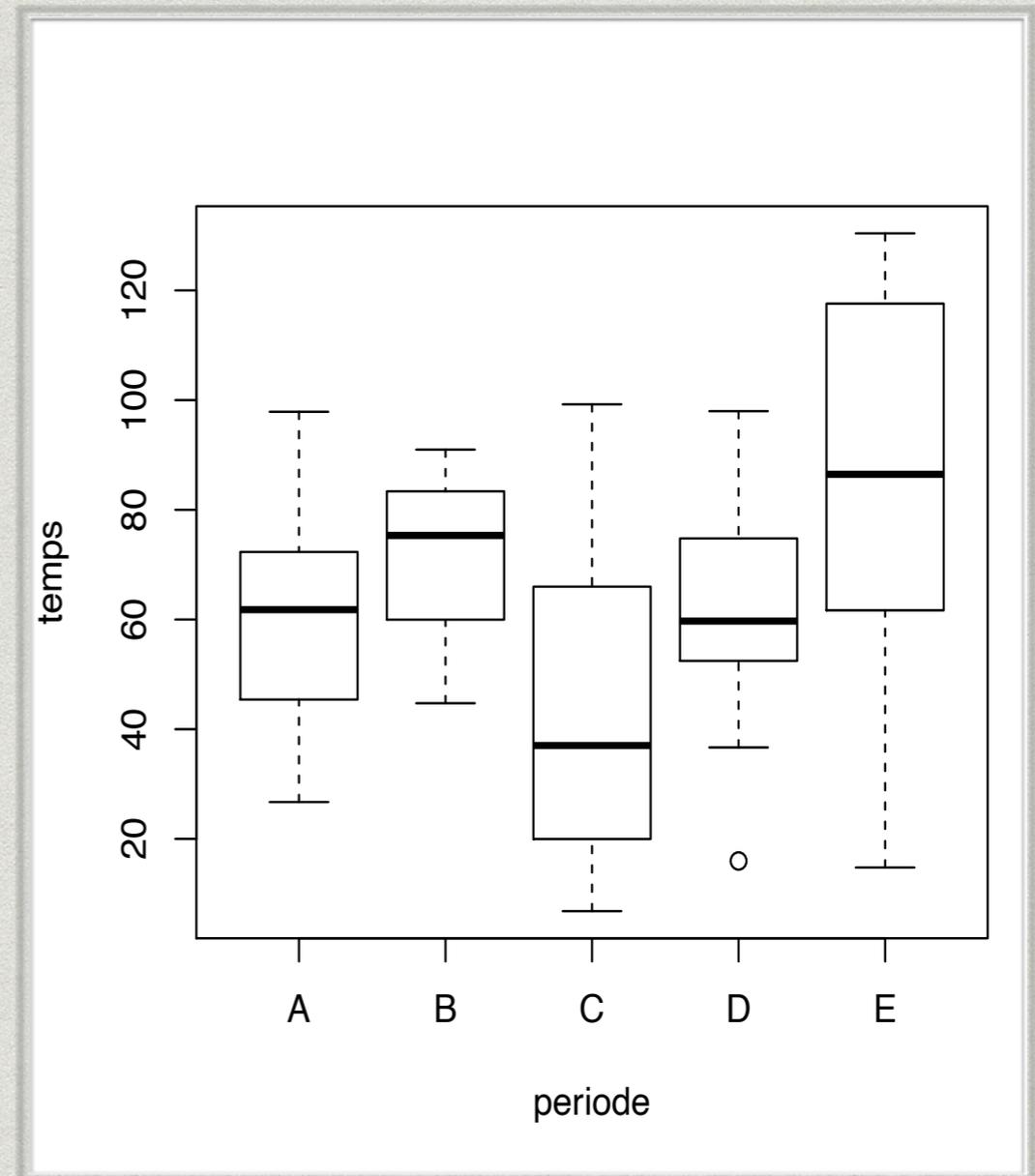
# Test du Khi-deux

- \* On s'intéresse aux intentions de vote aux élections présidentielles, *i.e.* le choix du candidat (**variable qualitative**) en interrogeant des étudiants d'une Ecole de Commerce à différentes années d'études (**variable qualitative**)
- \* Ici, le test statistique nous donne une p-value de 0.00005 ce qui nous conduit à rejeter l'hypothèse d'indépendance entre les deux variables



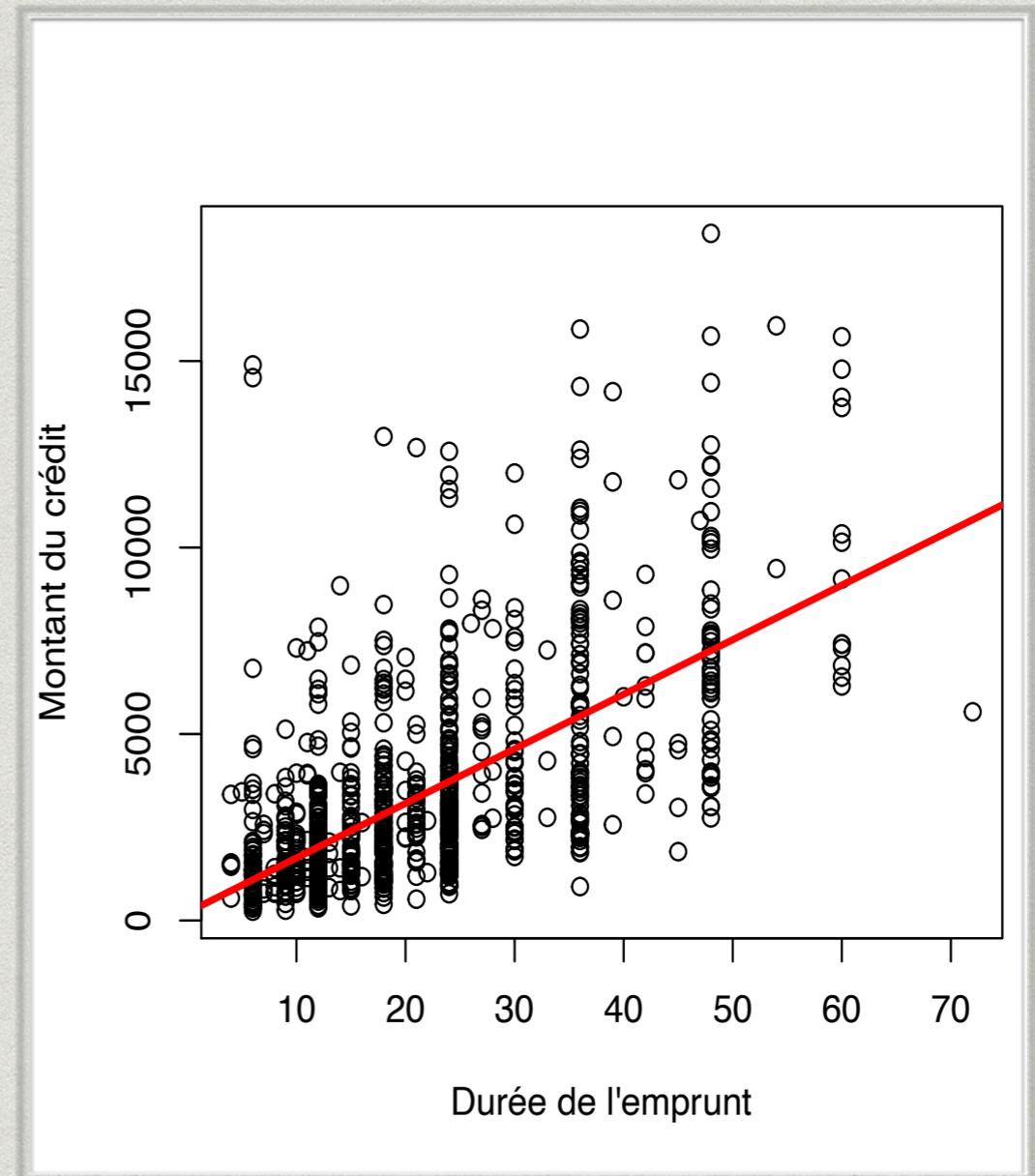
# Analyse de Variance

- \* Etude du temps passé à regarder à regarder la télévision (**variable quantitative**) en fonction de la catégorie d'âge des individus (**variable qualitative**)
- \* Ici, le test statistique nous donne une p-value de **0.00002** ce qui nous conduit à rejeter l'hypothèse d'indépendance entre les deux variables



# Régression linéaire

- \* Etude de la corrélation entre le montant du crédit emprunté (**variable quantitative**) et la durée de l'emprunt (**variable quantitative**).
- \* On enregistre a priori un coefficient de corrélation égale à **0.624**
- \* Un test de corrélation de Student retourne une p-value proche de 0. Donc le montant de l'emprunt est corrélée à sa durée.



# Théorie des tests

- \* Vous serez confronté à trois cas différents, *i.e.* trois tests statistiques différents selon la nature des variables
  1. Quali - quali : on procèdera à un test du  $\chi^2$
  2. Quanti - Quali : on effectue une analyse de variance (ANOVA)
  3. Quanti - Quanti : on parle de régression linéaire et on teste la corrélation avec un test de Student
- \* Avec Sphinx, vous n'aurez qu'une seule chose à faire, lire la p-value dans les deux premiers cas et lire la valeur de la statistique de test dans le cas 3.

# Lien avec l'analyse de données

- \* En statistiques, nous avons vu deux types d'*analyses*
  1. Les tris à plat : ils permettent de décrire la population selon différentes caractéristiques et permettent de dégager des tendances dans la population
  2. Les tris croisés : ils vont permettre de pousser l'analyse en étudiant les relations entre les variables comme dans les exemples vus précédemment
- \* Ces analyses se bases sur des données qui sont récoltées au cours d'une enquête.
- \* Les outils statistiques sont alors un *outil* permettant de dégager des informations, ensuite l'atout majeur ... c'est vous, votre esprit critique ainsi que les commentaires que vous pouvez formuler !

# EN PRATIQUE SOUS SPHINX

TRIS CROISÉS

**SE RENDRE SUR SPHINX**

**SÉLECTIONNER L'ENQUÊTE AUTOMOBILE**

**CLIQUER SUR ANALYSE**

**RETOUR À L'ENVIRONNEMENT  
STANDARD D'ANALYSE**

**ET SE RENDRE SUR TRIS CROISÉS**

Aperçu
Nouvelle analyse
DATAVIV<sup>®</sup> New!

Tri à plat
Tri croisé
Analyse textuelle
Analyses avancées

**Tableau**

Variable en ligne

22 - ENTRETIEN

Utiliser la mise en classe

Options

---

Variable en colonne

24 - ENTRETIEN2

Utiliser la mise en classe

Options

---

Analyses

Moyenne    Ecart-type    Médiane

Min-Max    Somme

Effectifs

Tests statistiques

---

Options de calcul

Ignorer les non-réponses

---

Affichage

Titre : Automatique

Afficher le tableau

Afficher le graphique

Afficher un commentaire personnalisé

## 22 - ENTRETIEN / 24 - ENTRETIEN2

|                   | Moyenne | Ecart-type | Effectif |
|-------------------|---------|------------|----------|
| <b>ENTRETIEN</b>  | 2,1     | 2,73       | 175      |
| <b>ENTRETIEN2</b> | 1132    | 812,74     | 175      |

**Réponses effectives : 175   Non-réponse(s) : 0   Taux de réponse : 100%**

**ENTRETIEN2 = 171 \* ENTRETIEN + 771**

EXEMPLE DE FENÊTRE QUE VOUS POURRIEZ AVOIR

TRI CROISÉ AVEC SPHINX

# Tris Croisés

- \* Choix de deux variables (d'où le terme **tri croisé** ou **analyse bivariée** !)
- \* Les menus restent globalement inchangés, et on conserve les mêmes options
- \* Ce qui va nous intéresser plus particulièrement ce sont **l'ordre des variables** dans la représentation et la petite icône « **Tests Statistiques** »

The screenshot shows a software interface for a cross-tabulation analysis. At the top, there are two tabs: "Tableau" and "Graphique". The interface is divided into several sections:

- Variante en ligne**: A dropdown menu shows "1 - possession". There are left and right arrow buttons. Below it is a checkbox for "Regrouper les modalités" and a link for "Options".
- Variante en colonne**: A dropdown menu shows "2 - MARQUE". There is an up arrow button. Below it is a checkbox for "Regrouper les modalités" and a link for "Options".
- Représentation**: Two radio buttons are present: "Tableau - Graphique" (selected) and "Carte AFC".
- Analyses**: Three checkboxes are shown: "Effectifs" (checked), "Pourcentages" (checked) with a dropdown menu set to "En ligne", and "Tests statistiques" (unchecked).
- Options de calcul**: A checkbox for "Ignorer les non-réponses" is shown.
- Affichage**: A "Titre" dropdown menu is set to "Automatique". Below it are three checkboxes: "Afficher le tableau" (checked), "Afficher le graphique" (checked), and "Afficher un commentaire personnalisé" (unchecked).

## 2 - MARQUE / 4 - APPARTENANCE

| MARQUE ↓      | APPARTENANCE → |              | A vous   |           | Une société de Votre |              | A quelqu'un de Votre famille |        | Total |        |
|---------------|----------------|--------------|----------|-----------|----------------------|--------------|------------------------------|--------|-------|--------|
|               | Eff.           | % Obs.       | Eff.     | % Obs.    | Eff.                 | % Obs.       | Eff.                         | % Obs. | Eff.  | % Obs. |
| Renault       | 29             | 80,6%        | 2        | 5,6%      | 5                    | 13,9%        | 36                           | 100%   |       |        |
| Peugeot       | 17             | 81%          | 2        | 9,5%      | 2                    | 9,5%         | 21                           | 100%   |       |        |
| Décret        | 9              | 75%          | 0        | 0%        | 3                    | 25%          | 12                           | 100%   |       |        |
| BMW           | 8              | 88,9%        | 0        | 0%        | 1                    | 11,1%        | 9                            | 100%   |       |        |
| Citroën       | 12             | 57,1%        | 3        | 14,3%     | 6                    | 28,6%        | 21                           | 100%   |       |        |
| Gué           | 13             | 92,9%        | 0        | 0%        | 1                    | 7,1%         | 14                           | 100%   |       |        |
| Mercedes-Benz | 7              | 87,5%        | 0        | 0%        | 1                    | 12,5%        | 8                            | 100%   |       |        |
| Opel          | 9              | 90%          | 0        | 0%        | 1                    | 10%          | 10                           | 100%   |       |        |
| Toyota        | 2              | 66,7%        | 0        | 0%        | 1                    | 33,3%        | 3                            | 100%   |       |        |
| Volvo         | 3              | 100%         | 0        | 0%        | 0                    | 0%           | 3                            | 100%   |       |        |
| Wolkswagen    | 13             | 72,2%        | 0        | 0%        | 5                    | 27,8%        | 18                           | 100%   |       |        |
| Autre         | 16             | 80%          | 0        | 0%        | 4                    | 20%          | 20                           | 100%   |       |        |
| <b>Total</b>  | <b>138</b>     | <b>78,9%</b> | <b>7</b> | <b>4%</b> | <b>30</b>            | <b>17,1%</b> | <b>175</b>                   |        |       |        |

Les pourcentages sont calculés par rapport au nombre d'observations en ligne.

Réponses effectives : 175

Non-réponse(s) : 0

Taux de réponse : 100%

# TRIS CROISÉS : ORDRE DES VARIABLES

## INFLUENCE DE L'ORDRE

EN LIGNE : MARQUE

EN COLONNE : APPARTENANCE

## 4 - APPARTENANCE / 2 - MARQUE

| APPARTENANCE ↓                      | Renault   |              | Peugeot   |            | Décret    |             | BMW      |             | Citroën   |            | Gué       |           | Mercedes-Benz |             | Opel      |             | Toyota   |             | Volvo    |             | Wolkswagen |              | Autre     |              | Total      |             |
|-------------------------------------|-----------|--------------|-----------|------------|-----------|-------------|----------|-------------|-----------|------------|-----------|-----------|---------------|-------------|-----------|-------------|----------|-------------|----------|-------------|------------|--------------|-----------|--------------|------------|-------------|
|                                     | Eff.      | % Obs.       | Eff.      | % Obs.     | Eff.      | % Obs.      | Eff.     | % Obs.      | Eff.      | % Obs.     | Eff.      | % Obs.    | Eff.          | % Obs.      | Eff.      | % Obs.      | Eff.     | % Obs.      | Eff.     | % Obs.      | Eff.       | % Obs.       | Eff.      | % Obs.       | Eff.       | % Obs.      |
| <b>A vous</b>                       | 29        | 21%          | 17        | 12,3%      | 9         | 6,5%        | 8        | 5,8%        | 12        | 8,7%       | 13        | 9,4%      | 7             | 5,1%        | 9         | 6,5%        | 2        | 1,4%        | 3        | 2,2%        | 13         | 9,4%         | 16        | 11,6%        | <b>138</b> | <b>100%</b> |
| <b>Une société de Votre</b>         | 2         | 28,6%        | 2         | 28,6%      | 0         | 0%          | 0        | 0%          | 3         | 42,9%      | 0         | 0%        | 0             | 0%          | 0         | 0%          | 0        | 0%          | 0        | 0%          | 0          | 0%           | 0         | 0%           | <b>7</b>   | <b>100%</b> |
| <b>A quelqu'un de Votre famille</b> | 5         | 16,7%        | 2         | 6,7%       | 3         | 10%         | 1        | 3,3%        | 6         | 20%        | 1         | 3,3%      | 1             | 3,3%        | 1         | 3,3%        | 1        | 3,3%        | 0        | 0%          | 5          | 16,7%        | 4         | 13,3%        | <b>30</b>  | <b>100%</b> |
| <b>Total</b>                        | <b>36</b> | <b>20,6%</b> | <b>21</b> | <b>12%</b> | <b>12</b> | <b>6,9%</b> | <b>9</b> | <b>5,1%</b> | <b>21</b> | <b>12%</b> | <b>14</b> | <b>8%</b> | <b>8</b>      | <b>4,6%</b> | <b>10</b> | <b>5,7%</b> | <b>3</b> | <b>1,7%</b> | <b>3</b> | <b>1,7%</b> | <b>18</b>  | <b>10,3%</b> | <b>20</b> | <b>11,4%</b> | <b>175</b> |             |

Les pourcentages sont calculés par rapport au nombre d'observations en ligne.

Réponses effectives : 175

Non-réponse(s) : 0

Taux de réponse : 100%

# TRIS CROISÉS : ORDRE DES VARIABLES INFLUENCE DE L'ORDRE

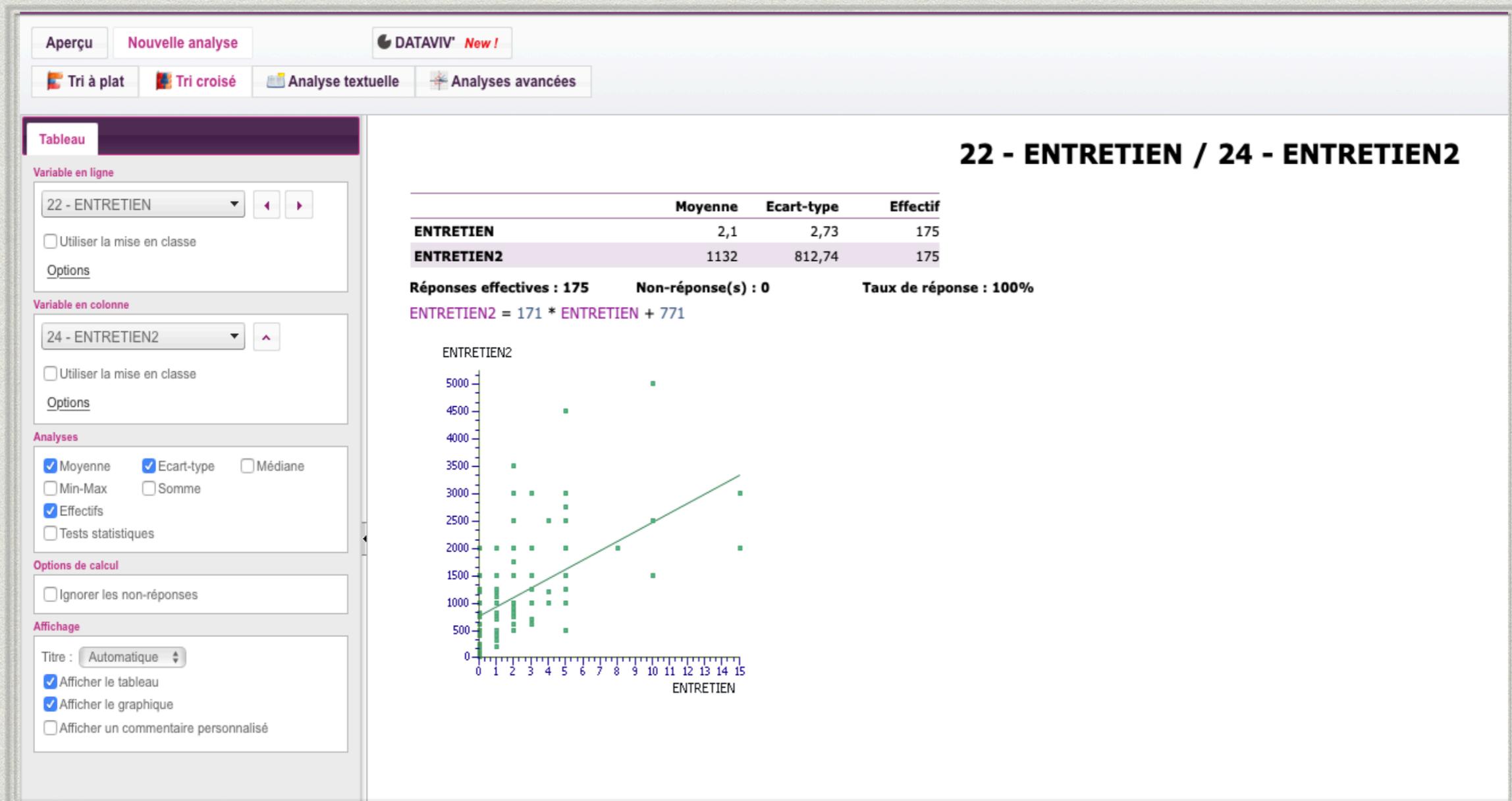
EN LIGNE : APPARTENANCE

EN COLONNE : MARQUE

# Tris Croisés : Remarques

- \* Notes les différences au niveau des pourcentages indiqués par Sphinx selon l'ordre des variables
- \* Prenez l'exemple au croisement entre « Renault » et « A vous ». Dans les deux cas, on trouve que 29 répondants possèdent une Renault. Sauf que dans le premier tableau, ces 29 répondants correspondent à 80,6% des propriétaires d'une Renault (**variable en ligne Marque**) alors que, dans le deuxième cas, cela correspond à 21% des propriétaires d'un véhicule (**variable en ligne Appartenance**)
- \* Vous remarquez que les pourcentages de chaque ligne somment à 100% contrairement aux colonnes. Faites donc bien attention aux choix des variables en *ligne* et *colonne* pour la lecture de votre tableau.

# Tris Croisés : Quanti - Quanti Entretien vs Entretien2



# Tris Croisés : Quanti - Quanti

## Entretien vs Entretien2

- \* On teste la corrélation entre deux variables quantitatives.
- \* On effectue donc une **régression linéaire** et un **test de corrélation de Pearson** basé sur la loi de Student.
- \* Sphinx vous donne un **modèle de corrélation** entre les deux variables, *i.e.* **la droite qui passe par le maximum de points** et qui minimise une certaine erreur (une erreur quadratique, mais cela n'a pas d'importance ici).
- \* Cliquez maintenant sur la case « Tests Statistiques »

# Tris Croisés : Quanti - Quanti

## Entretien vs Entretien2

|                   | Moyenne     | Ecart-type | Effectif |
|-------------------|-------------|------------|----------|
| <b>ENTRETIEN</b>  | <b>2,1</b>  | 2,73       | 175      |
| <b>ENTRETIEN2</b> | <b>1132</b> | 812,74     | 175      |

Les valeurs en bleu / rouge sont significativement supérieures / inférieures à la grande moyenne (au seuil de risque de 5%).

Réponses effectives : 175      Non-réponse(s) : 0      Taux de réponse : 100%  
Corrélation = 0,58. L'ajustement entre les deux variables est plutôt faible ( $t = 9,27$ ).

$$\text{ENTRETIEN2} = 171 * \text{ENTRETIEN} + 771$$

- \* Affichage du coefficient de corrélation 0,58 : donc une corrélation plutôt moyenne
- \* La valeur de la statistique de **test de Student**  $t = 9,27$  ainsi que la moyenne et l'écart-type de chaque variable (utile pour calculer la statistique de test).
- \* Pas de p-value ! Mais on retiendra que l'on rejette l'hypothèse d'indépendance lorsque  $|t| > 2$
- \* On peut donc affirmer que les deux variables sont corrélées dans le cas présent (Remarque changer l'ordre des variables ne change pas la conclusion)

# Tris Croisés : Quali - Quali Marque vs Appartenance

## 2 - MARQUE / 4 - APPARTENANCE

| MARQUE ↓      | APPARTENANCE → |              |       | A vous   |           |       | Une société de Votre |              |       | A quelqu'un de Votre famille |        |       | Total |        |
|---------------|----------------|--------------|-------|----------|-----------|-------|----------------------|--------------|-------|------------------------------|--------|-------|-------|--------|
|               | Eff.           | % Obs.       | Ecart | Eff.     | % Obs.    | Ecart | Eff.                 | % Obs.       | Ecart | Eff.                         | % Obs. | Ecart | Eff.  | % Obs. |
| Renault       | 29             | 80,6%        |       | 2        | 5,6%      |       | 5                    | 13,9%        |       | 36                           | 100%   |       |       |        |
| Peugeot       | 17             | 81%          |       | 2        | 9,5%      |       | 2                    | 9,5%         |       | 21                           | 100%   |       |       |        |
| Décret        | 9              | 75%          |       | 0        | 0%        |       | 3                    | 25%          |       | 12                           | 100%   |       |       |        |
| BMW           | 8              | 88,9%        |       | 0        | 0%        |       | 1                    | 11,1%        |       | 9                            | 100%   |       |       |        |
| Citroën       | 12             | 57,1%        | - TS  | 3        | 14,3%     | + S   | 6                    | 28,6%        | + PS  | 21                           | 100%   |       |       |        |
| Gué           | 13             | 92,9%        |       | 0        | 0%        |       | 1                    | 7,1%         |       | 14                           | 100%   |       |       |        |
| Mercedes-Benz | 7              | 87,5%        |       | 0        | 0%        |       | 1                    | 12,5%        |       | 8                            | 100%   |       |       |        |
| Opel          | 9              | 90%          |       | 0        | 0%        |       | 1                    | 10%          |       | 10                           | 100%   |       |       |        |
| Toyota        | 2              | 66,7%        |       | 0        | 0%        |       | 1                    | 33,3%        |       | 3                            | 100%   |       |       |        |
| Volvo         | 3              | 100%         |       | 0        | 0%        |       | 0                    | 0%           |       | 3                            | 100%   |       |       |        |
| Wolkswagen    | 13             | 72,2%        |       | 0        | 0%        |       | 5                    | 27,8%        |       | 18                           | 100%   |       |       |        |
| Autre         | 16             | 80%          |       | 0        | 0%        |       | 4                    | 20%          |       | 20                           | 100%   |       |       |        |
| <b>Total</b>  | <b>138</b>     | <b>78,9%</b> |       | <b>7</b> | <b>4%</b> |       | <b>30</b>            | <b>17,1%</b> |       | <b>175</b>                   |        |       |       |        |

Les pourcentages sont calculés par rapport au nombre d'observations en ligne.

Les valeurs en bleu / rouge sont significativement sur représentées / sous représentées (au seuil de risque de 5%).

Réponses effectives : 175      Non-réponse(s) : 0      Taux de réponse : 100%  
p-value = 0,57 ; Khi2 = 20,21 ; ddl = 22,00. La relation n'est pas significative.

# Tris Croisés : Quali - Quali Marque vs Appartenance

**Réponses effectives : 175      Non-réponse(s) : 0      Taux de réponse : 100%**  
**p-value = 0,57 ; Khi2 = 20,21 ; ddl = 22,00. La relation n'est pas significative.**

- \* On cherche à étudier le lien éventuel entre l'appartenance du véhicule (**variable qualitative**) et la marque du véhicule (**variable qualitative**)
- \* Le test effectué est un **test du Khi-deux** à 22 degrés de liberté (compter le nombre de modalité de chaque variable)
- \* La statistique de test est égale à 20,21 sur notre échantillon ce qui correspond à une **p-value de 0,57**
- \* La p-value étant plus grande que le risque de première espèce  $\alpha$ , rien ne permet de rejeter l'hypothèse nulle  $H_0$ , on ne peut conclure au rejet de l'indépendance des deux variables. Donc cette hypothèse reste vraie

# Tris Croisés : Quanti - Quali

## Entretien2 vs Puissance

| 24 - ENTRETIEN2 / 7 - PUISSANCE |             |             |           |          |          |          |               |        |
|---------------------------------|-------------|-------------|-----------|----------|----------|----------|---------------|--------|
| ENTRETIEN2 ↓                    | PUISSANCE → | Non-réponse | - De 4 CV | 4 à 5 CV | 6 à 7 CV | 8 à 9 CV | 10 CV et plus | Total  |
| Moyenne                         |             | 785,29      | 575       | 824,29   | 1256,38  | 1941,67  | 2900          | 1132   |
| Ecart-type                      |             | 781,38      | 237,89    | 552,58   | 569,86   | 929,08   | 1024,7        | 812,74 |
| Effectif                        |             | 17          | 12        | 70       | 47       | 24       | 5             | 175    |

Réponses effectives : 158      Non-réponse(s) : 17      Taux de réponse : 90,3%

- \* On cherche étudier une éventuelle corrélation entre le coût de l'entretien2 du véhicule (**variable quantitative**) et la puissance du véhicule (**variable qualitative**) (notez l'utilisation de la mise en classe pour cette variable)
- \* Pour procéder à l'analyse d'éventuelles corrélations entre ces deux variables, on va procéder à **une analyse de variance (ANOVA)**
- \* Cliquez sur l'icône « Tests Statistiques » pour voir ce qu'il en est

# Tris Croisés : Quanti - Quali

## Entretien2 vs Puissance

### 24 - ENTRETIEN2 / 7 - PUISSANCE

|                     | Non-réponse | - De 4 CV  | 4 à 5 CV      | 6 à 7 CV | 8 à 9 CV       | 10 CV et plus | Total         |
|---------------------|-------------|------------|---------------|----------|----------------|---------------|---------------|
| <b>ENTRETIEN2 ↓</b> |             |            |               |          |                |               |               |
| <b>Moyenne</b>      | 785,29      | <b>575</b> | <b>824,29</b> | 1256,38  | <b>1941,67</b> | <b>2900</b>   | <b>1132</b>   |
| <b>Ecart-type</b>   | 781,38      | 237,89     | 552,58        | 569,86   | 929,08         | 1024,7        | <b>812,74</b> |
| <b>Effectif</b>     | 17          | 12         | 70            | 47       | 24             | 5             | <b>175</b>    |

Les valeurs en bleu / rouge sont significativement supérieures / inférieures à la grande moyenne (au seuil de risque de 5%).

Réponses effectives : 158      Non-réponse(s) : 17      Taux de réponse : 90,3%  
 p-value = < 0,01 ; Fisher = 21,34. La relation est très significative.

- \* On rappelle que l'hypothèse nulle  $H_0$  consiste à supposer que les variables sont indépendantes
- \* L'analyse de variance nous retourne une statistique de Fisher égale à 21,34, ce qui correspond à une p-value plus petite que 0,01.
- \* Le risque de première espèce  $\alpha$  étant généralement fixé à 5%, on en conclut donc au rejet de l'hypothèse  $H_0$  et donc à la corrélation entre les variables *Entretien2* et *Puissance*.
- \* L'analyse du tableau nous montre même que les véhicules les plus puissances sont ceux dont le coût de l'entretien est le plus élevé contrairement aux véhicules les moins puissants. Il faut imaginer que l'usure des pièces pour ce type de véhicule est beaucoup plus rapide.

# Quelques remarques

- \* Il est possible de changer les échelles en nombre, notamment lorsque l'on étudie une note (**variable qualitative**). Le passage en nombre (**variable quantitative**) peut parfois s'avérer nécessaire, pour le calcul d'une note moyenne par exemple. Mais il va surtout **changer la nature de la variable et donc du test statistique employé.**
- \* Pour ce qui va suivre, gardez à l'esprit que vos analyses et commentaires sont importants pour votre travail !

# **TRAVAIL DE SYNTHÈSE EN GROUPE ET DERNIER QUIZZ DE CE COURS**

**DIRECTION BRIGHTSPACE ET LISEZ LES CONSIGNES**

**LE TRAVAIL S'EFFECTUERA PAR GROUPE DE 2 OU 3 PERSONNES**